

本周周报（2013.11.25-2013.12.01）

郭方舟

本周工作

1 空气污染数据可视化

本周大部分的时间都在讨论项目下一步的方向。

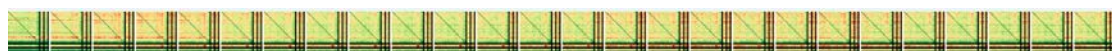
首先，跟斐然师兄讨论了转移熵的含义。

延时为 1 的序列 J 到序列 I 的转移熵计算公式如下：

$$T_{J \rightarrow I} = \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) \log \left(\frac{p(x_{n+1}, x_n, y_n) \cdot p(x_n)}{p(x_n, y_n) \cdot p(x_{n+1}, x_n)} \right)$$

从这个式子中可以看到，如果 x_n 恒等于 y_n 的话，那么转移熵就等于 0；我们猜想如果 y_n 恒等于 x_{n+1} 的话，转移熵取最大。之后我们做了几个例子进行实验，当序列 I 为序列 J 后移延时 k 的时候，序列 J 到序列 I 的转移熵在延时为 k 的时候取到最大。因此我们认为转移熵在表达序列 J 对序列 I 的影响时，实际是刻画 $y_n = x_{n+k}$ 的数量。

之后的讨论则是关于转移熵怎么使用，我们进行了加窗，进行窗移计算转移熵，但是发现这样算出来的转移熵的区别并不大。反而计算延时 1-n 的转移熵后发现了一些周期性的规律：



从上图可以看到，转移熵矩阵整体上颜色是有一定的周期性变化的。

周日的讨论则提到在提取 motif 之后将 motif 视为文本中的词，并将

整个数据当成文本进行主题抽取，这样的方法最理想的结果是能够找出类似于下雨，下雪等等的主题。

2 课程作业

基本完成了体系结构的综述。

开始动手完成网格计算的作业。

12 月 21 号要考托福，因此这个礼拜花了一些时间复习托福。

下周工作

1. 空气污染数据的可视化

尝试生成 motif，并进行主题抽取。

2. 作业

完成网格计算的作业&考托福。